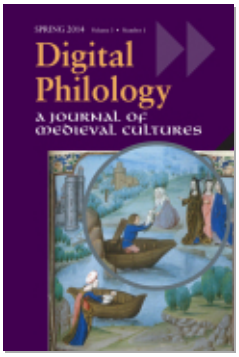
An Electronic Corpus of Fifteenth-Century Castilian Cancionero Manuscripts



Barbara Bordalejo, Fiona Maguire, Manuel Moreno, Peter Robinson, Dorothy Severin

Digital Philology: A Journal of Medieval Cultures, Volume 3, Number 1, Spring 2014, pp. 11-23 (Article)

Published by The Johns Hopkins University Press
DOI: 10.1353/dph.2014.0003

Barbara Bordalejo
University of Birmingham

Fiona Maguire
University of Liverpool

Manuel Moreno
University of Liverpool

Peter Robinson
University of Saskatchewan

Dorothy Severin
University of Liverpool

An Electronic Corpus of Fifteenth-Century Castilian *Cancionero* Manuscripts

This article describes the project *An Electronic Corpus of Fifteenth-Century Castilian* Cancionero *Manuscripts*, paying particular attention to issues and solutions in matters related to textual criticism, such as the collation of witnesses and the generation of phylogenetic stemmata.

> "L'écrit électronique…reproduit l'œuvre médiévale dans sa variance même."
>
> Bernard Cerquiglini, *Éloge de la variante* 116

1. Overview

The digital library *An Electronic Corpus of Fifteenth-Century Castilian* Cancionero *Manuscripts* is a free, open-access website which makes available the large corpus of fifteenth-century (c. 1360–1520) Castilian *cancionero* poetry, comprising over 4,000 individual poems extant in over 150 manuscript witnesses and over 150 early printed sources, conserved in archives in Spain, England, France, Italy, Portugal, Switzer-

land, and the USA. The website provides codicological MS descriptions, digitized MS images and digital transcriptions of the corpus, including variorum editions of poems that occur in multiple witnesses, with their collations and phylogenetic stemmata displayed. The website incorporates a digital edition of the late Brian Dutton's seminal seven-volume selected edition of the corpus, *El cancionero del siglo XV, 1360–1520*, born digital but published hitherto in printed form only.

1.1 Project Team

The digital library has been produced through an Arts and Humanities Research Council of Great Britain (AHRC)-funded research project collaboration between the Universities of Liverpool and Birmingham, UK, and Barcelona, Spain, comprising the following project team members: project director Dorothy Severin (University of Liverpool); co-directors Peter Robinson (formerly University of Birmingham, UK, now University of Saskatchewan, Canada), Vicenç Beltran (formerly University of Barcelona, now Università degli Studi di Roma "La Sapienza," Rome, Italy); research associates: Barbara Bordalejo (University of Birmingham), Manuel Moreno (University of Liverpool) and Fiona Maguire (University of Liverpool); technical assistance: Duncan Appelbe and Jake Gannon (University of Liverpool), Andrew West (University of Birmingham). Additional contributors include the late Elena Carrillo (Universiteit Utrecht, Holland), María-Jesús Díez Garretas (Universidad de Valladolid, Spain), and Andrea Zinato (Università di Verona, Italy).

1.2 Financial Support

The project has been financed in consecutive phases by the Arts and Humanities Research Council of GB (AHRC) (2004–07), the Modern Humanities Research Association (MHRA) (2008–10), and the Leverhulme Trust (2011–13). Financial and technical support has also been provided by the University of Liverpool.

2. Description of the Fifteenth-Century Castilian *Cancionero* Corpus

The very extensive corpus of fifteenth-century Castilian *cancionero* manuscripts comprises works composed in Castilian between c.1360–1520. The hetereogeneity of this corpus, containing mainly poetry, but also some prose works, is reflected in: the varied types of codices—anthologies, miscellanies; single author, multiple author collections—production in both Peninsular Spanish kingdoms and the Neapolitan-Aragonese kingdom; diverse literary themes (courtly, moral-didactic, political, reli-

gious, satirical, and occasional verse) and poetic genres (*canción*, *decir*, *soneto*, *pregunta*, *respuesta*, *glosa*); paratextual prose; and verse and prose treatises (*Tratados*). The corpus includes both non-musical and musical *cancioneros*.

3. Contents of *Electronic Corpus of "Cancionero" Manuscripts* Website

3.1 Catalogue and Indices

The *cancionero* digital website produced by the project adopts the standard catalogue and index reference system for Castilian *cancionero* poetry and *cancionero* witnesses, devised and first published by Brian Dutton in his *Catálogo-Índice de la poesía cancioneril del siglo XV* (1982), with addenda in his *El cancionero del siglo XV* (1990–91). The Dutton system assigns:

a) an ID number to each independent work, or item (whether poem or prose) per *cancionero* witness; if the same item is extant in other witnesses, the item bears the same ID number; relationships between items are also reflected in the ID numbering system: e.g. [ID0073 S 0072] denotes item ID0073 occurs in a series (indicated as S) beginning with item ID0072; [ID0030 G 0029] denotes item 0030 is a gloss of item 0029; [ID0091 P 0050] denotes item 0091 is a prose introduction or prologue to item 0050; (for a complete list of Dutton conventions see Dutton, *El cancionero* 1: vii).

b) a letter to MS witnesses, denoting archive location, followed by a number: e.g. LB3 = **L**ondon, **B**ritish Library, **3** (one of three fifteenth-century Castilian *cancioneros* held in the British Library). London, BL, MS [Egerton 939], *Cancionero de Egerton*, 122 ff.

c) a number to printed sources, taking the last two digits of the year of publication, preceded by an asterisk if the date is uncertain, followed by letters related to the author or name of the work, or printed work: e.g. 11CG = ***C**ancionero **g**eneral* (Valencia, 15**11**); 04MC = **M**ena (Juan de), *La **C**oronación* (Toledo, 15**04**).

The Dutton ID system facilitates not only the identification of a work within a *cancionero* witness, but also easy identification of the same item in other witnesses and the order of items within witnesses.

Fig. 1: Collation View, [ID0028] Stanza 1, line 3, Santillana, *El Infierno de los enamorados*.

3.2 Codicological MS Descriptions

Full codicological descriptions of some of the fifteenth-century MS witnesses are available on the website; the rest are in preparation, forthcoming on the website during 2014.

3.3 MS Digital Images

Digital images of *cancionero* MSS in this corpus are being made available on the website where publication permission from archives has been granted and/or links to archives' own digital collections exist, for example: *Cancionero de Baena* [BNF, MS espagnol 37], Dutton PN1, digital images available at BnF.[1] Gallica

3.4 Transcriptions

Digital transcriptions of the poetic corpus are available for the vast majority of the poems in the Castilian *cancioneros* in each of their manuscript witnesses, and early printed witnesses in some cases (see 4.1 below).

3.5 Collation

Poems extant in multiple witnesses transcribed by Severin and Maguire have been collated electronically, and results displayed in the Collation View window (Fig. 1); for a discussion of the methodology used to create digital text collation, see 4.2 below.

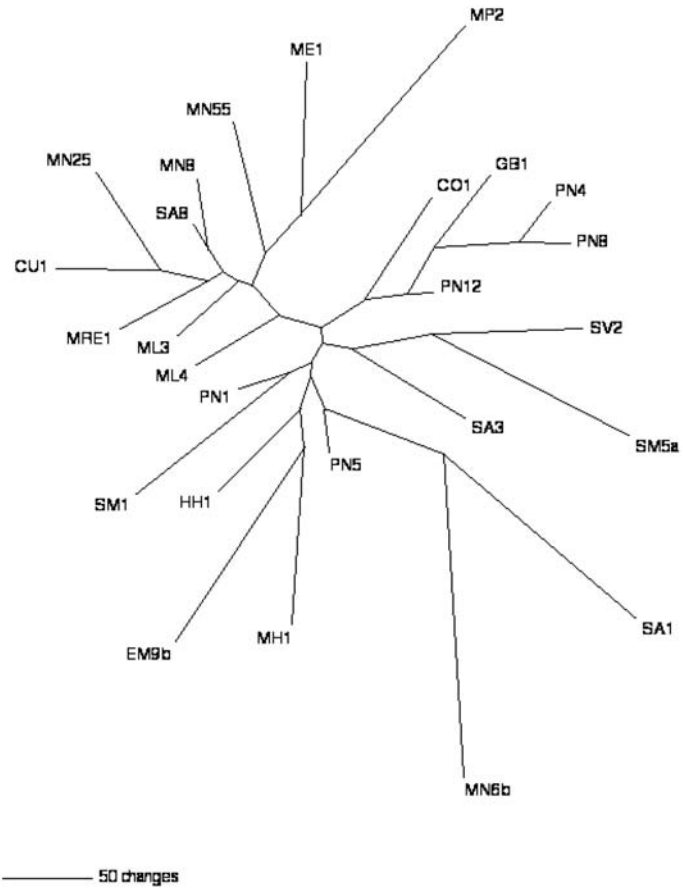


Fig. 2: Phylogenetic Stemma, Unrooted Phylogram of ID0050, Santillana, *Proverbios* (generated from variorum edition by Elena Carrillo).

3.6 Phylogenetic Stemmata

Phylogenetic stemmata, generated by software developed for evolutionary biology research, have been created for variorum editions of poems (Fig. 2) and are accessible in Collation View (see 4.3 below).

3.7 Consultation of Editions

Users can consult the Severin-Maguire editions in three distinct windows: Transcription View, Text/Image View, and Collation View; phylogenetic stemmata of variorum editions are accessible in Collation View. Poems in multiple witnesses in the Dutton editions can be compared by drawing the mouse over the texts.

4. Methodology

4.1 Transcriptions

The transcriptions available on the *cancionero* website were all born digital and derive from two sources:

4.1.1. Dutton Editions

Brian Dutton's selected transcriptions of the corpus, originally published in printed book form as *El cancionero del siglo XV* (1990–91) but born digital, were converted from ASCII files to XML. The Dutton editions are semi-diplomatic, with expansion of scribal abbreviations displayed in italics.

4.1.2 Severin-Maguire Editions

Transcriptions of the remainder of the corpus have been prepared by Severin and Maguire following different norms of transcription: the editions are diplomatic, abbreviations have been encoded using Junicode font, and TEI-conformant tags have been used to indicate emendations to the codices, whether by the original scribal hand or other hands.

4.2 Collation

4.2.1 Dutton Editions

The Dutton editions have not been collated electronically to date; poems extant in multiple witnesses can be compared by moving the mouse over the text.

4.2.2 Severin-Maguire Editions

Poems extant in multiple witnesses transcribed by Severin-Maguire have been collated electronically by the editors using Collate 2.0 software (software developed by project co-director Peter Robinson). The collated digital data results per poem ID are displayed for the user in the Collation View window, showing significant variants together with the original orthographic form for each witness per word, line, and stanza.

In her article "The Transcription and Encoding of pre-1520 *Cancionero* Manuscripts," available on the project website, Bordalejo describes the transcription and collation process as follows: in order to compare different texts, Collate requires a base text for collation which is used as a reference to establish what is present or absent in every witness. Several criteria can be used to decide how to choose a base text for collation but, generally speaking, it has been simpler to create an artificial one (based on a witness chosen by the collator and "completed"

with all lines present in all witnesses). This has been the process used for the Severin-Maguire Corpus, where every single line of text is collated, making it possible to establish whether the witnesses present information that is unique, restricted to a part of the textual tradition or almost invariably present. This also allows the editor to determine whether there are changes in line or stanza order, and whether such changes are stemmatically significant.

The Severin-Maguire Corpus is original work and, for this reason, under our control. It was possible for us to assign the lineation in a meaningful manner. For example, in a case in which a stanza appears in some versions and not in others, we could assign stanza numbers to show this or to show the re-arrangement of lines within a stanza. If we had not proceeded in this manner then changes in the order of lines or stanzas could not have been detected. Moreover, the whole collation process would have been compromised, since lines or stanzas that do not correspond to each other would have appeared as equivalent.

4.2.2.1 The Lineation System

The lineation system for each particular ID follows the line order of the text chosen as a base for collation. This means that any diversion on the order of lines in witnesses other than the base will appear as disorder. Because this can have consequences from an interpretive perspective, it is important to take it into consideration when choosing a base for collation. However, in a project as vast as this one, it might not always be feasible to look at all the witnesses before deciding which one to use as a base for collation, in such cases, the best option is to warn the reader that the lineation system is arbitrary and does not intend to represent the character of line or stanzas.

Any order or disorder of stanzas and lines is more likely to be the result of editorial judgment than of an intrinsic value of a particular version over another. As stated above, it was necessary to proceed in this manner in order to be able to carry out the collation of the texts in a meaningful way—there is no point in comparing lines that are in the same relative position but are clearly not textually correspondent.

4.2.2.2 Additional Lines, Repeated Lines

Lines not present in the base text are tagged as "additional" lines, which means that they have been given a different number from the lines present in the base text. Generally, the lines in the base text are numbered consecutively, and lines not found in the base text, but found in other manuscripts are numbered "-1," "-2," "-3," etc. Occasionally, a manu-

script repeats a line. This generally happens when a scribe finishes one folio with a particular line and then starts the next folio with the same one. Alternatively, a manuscript might present the same line or stanza repeated many folios apart. When repeated lines are present, these need to be named separately from the original line (or they cannot be part of the collation) and they must be included in the base text. In this way, their presence and position can be accounted for in reference to all witnesses. To mark a repeated line, the letter *r* is added at the end of the line number.

4.2.2.3 Character Set

Scribal abbreviations in the Severin-Maguire editions have been encoded using Junicode, a Unicode font for medievalists. Users of the *cancionero* website should view the transcriptions using the Firefox browser. For examples of individual characters and scribal abbreviations see Bordalejo, "The Transcription and Encoding." For an example of MS witness transcription with scribal abbreviations see Santillana, *El infierno* [Dutton ID 0028], Madrid, BNE, MS 22.335 [Dutton MN55], fol. 38r.[2]

4.2.2.4 The Tagging System

The tagging system used for this project is based on the one originally created for the transcriptions of Dante's *Commedia*. Before we developed this system, similar projects attempted to present simultaneously both "what is in the manuscript" as a series of additions or deletions, and "what is in the text" as a series of distinct, discreet readings. However, discussions with Klaus Wachtel (Institut für Neutestamentliche Textforschung, Westfälische Wilhelms-Universität, Münster) about the transcription of corrections of the manuscripts of the Greek New Testament, generated new ideas about how to encode different textual stages. The main goal of the new transcription system is to present a clear distinction between what is in the manuscript and how the transcriber (or the editor) interprets the different stages of development of the text. These two levels should be clearly distinguished. Although both of them are interpretive, the first attempts to show what appears to be present in a manuscript and the second offers an opinion which informs what the editor thinks is going on at a particular point.

4.2.2.5 Places of Variation

The Dante *Commedia* Project encoding system aims to represent the different stages of variation in the text. When a transcriber finds a "place of variation" in the manuscript, he or she can use the "apparatus tag." The apparatus tag contains three main components: the original read-

ing, the final reading (contained in a tag which specifies which copyist produced it), and what "literally" is in the manuscript. If there are more than two stages in a correction, for example, in the case of having more than one corrector, these stages are presented in what is likely to be their successive order. The following example is taken from MS Modena, Bibl. Estense, MS [α.R.8.9], [Dutton ME1], fol. 63r, stanza 1, line 6.[3] We encode this – penas > pena – in the following manner:

[app][orig]penas[/orig][c1]pena[/c1][lit]pena[ud]s[/ud][/lit][/app]

What this means is that we know that the scribe originally wrote the word "penas"; subsequently, the final *s* was deleted by means of an expunction mark. Following this reasoning, the final reading intended by the scribe is "pena." To make it even clearer, we also include a [lit] tag we have named "literal" tag, in which we attempt to explain what we see in the document. Within the literal tag we include the under-dotted letter and we mark it as such. In this way, the reader's attention is called to that particular place of variation and s/he can refer to the image before deciding whether s/he agrees with our judgement.

Here is another example, taken from Paris, BN, MS Esp. 230, [Dutton PN8], folio 39r, stanza 3, line 1.[4] We encode this in the following manner:

[app][orig]gracia[/orig][c2]grecia[/c2][lit]gr[rp][cow]a[/cow][chk]e[/chk][/rp]cia[/lit][/app]

In this example, we see a letter that has not been erased, but that has been rewritten. As before, the original reading is stated, as well as the final one. Within the literal tag, we find that the letter *e* has been written on top of the letter *a*. The separation of the different stages of correction is useful because they allow us to have more clearly defined boundaries between what we believe to be different stages of the text. The system could be useful to understand scribal behaviour (by evaluating the changes a particular scribe is likely to make in a manuscript) and might allow us to see further layers of what would otherwise be just interpreted as conflation. That is, if a particular manuscript, after having been copied, happened to be corrected against a witness of a different recension, we should be able to assess the situation by analysis of the second stage of correction.

The advantages over systems that overlook the separation of stages of correction are too important to be easily dismissed. Firstly, the tran-

scribers can defer interpretation of the stages of meaning, since the literal tag can be transcribed independently of the other components of the apparatus tag (this also gives the advantage of allowing the editor of a publication to make a final decision as to what happened at each individual place of variation). Secondly, the content of the literal tag allows us to reconstruct what actually appears in a manuscript on the computer screen. Thirdly, the other components of the apparatus tag, such as original reading, final reading, and intermediate readings, can be collated separately from the rest of the text. The separate collation of multiple readings in a manuscript will be most useful when a scribe used a manuscript of different affiliation to correct his copy. In such cases, separate collation will allow the isolation of readings which originated in different manuscripts and which could hint at distinct affiliations in a single text. Separate collation might also be of help in cases in which conflation has occurred because a manuscript is corrected with another one from a different branch of a textual tradition.

This encoding system has also been implemented by *The Canterbury Tales Project* (*CTP*) for publications to appear after the *Miller's Tale* on CD-ROM, edited by Peter Robinson, and *The Nun's Priest's Tale* on CD-ROM, edited by Paul Thomas.[5] This encoding system could also offer advantages when applied to authorial manuscripts, and although it was originally designed to deal with problems of corrections presented by medieval manuscripts, it should work as efficiently to distinguish different authorial versions of a particular text. This should translate into an easier reconstruction of these versions, and allow the distinction and separate reconstruction of different authorial stages of composition, thus permitting the creation of genetic editions.

4.3 Phylogenetic Stemmata

Project co-director Peter Robinson has pioneered the application of new methods to stemmatics. Robinson, working with evolutionary biologists, first applied the computer techniques of evolutionary biology research in relation to the transmission of large Chaucer manuscript traditions in "The Phylogeny of *The Canterbury Tales*" (Barbrook, Howe, Blake, and Robinson).[6] The *cancionero* project has similarly applied these phylogenetic methods to longer *cancionero* poems which are extant in more than four witnesses. The digital data generated from the collation using the Collate programme is processed using PAUP 4.0 software. An explanation of the techniques and results, "About These Stemmata" by Peter Robinson, is available on the website alongside the stemmata displayed. Robinson observes:

> Considerable experimentation has been carried out in the last decades on the use of evolutionary biology programs, designed for the creation of hypotheses of genetic relationships among living objects, based on analysis of the characters they share and do not share. It has been found that the same programs can be used for the creation of hypotheses of genetic relationships among manuscripts….

Two types of phylogentic stemmata are displayed:

4.3.1 Unrooted Phylogram

This presents all the manuscripts in relation to one another without any presumption as to which is the archetype, or might be nearest to the archetype, and with an indication of how close manuscripts are to one another. Broadly, manuscripts descended from the same node are (the table suggests) related […]. (Robinson, "About These Stemmata")

4.3.2 Bootstrap Cladogram

This presents all the manuscripts in relation to one another without any presumption as to which manuscript is the archetype, and with an indication of how reliable each grouping of manuscripts is. The reliability of the grouping is indicated by the number (a percentage) on each line…. (Robinson, "About These Stemmata").

4.4 Manuscript Digital Images

Digital images of manuscript witnesses are in JPEG format and can be viewed from the MS Images page as well as in Text/Image view, alongside transcriptions.

5. Sustainability

The website is maintained by the University of Liverpool Computing Services Dept.

Conclusion

Launched in 2007, development of this digital corpus of fifteenth-century Castilian *cancionero* manuscripts website currently remains ongoing: work is in progress to add further variorum editions with collations and phylogenetic stemmata, and codicological MS descriptions, towards completion of the digital *cancionero* portal in 2014–15.

Notes

1. <http://tinyurl.com/ltdvbgp>
2. <http://tinyurl.com/lfboqls>.
3. <http://tinyurl.com/md9pax3>.
4. <http://tinyurl.com/nx9c78n>.
5. For further information on *The Miller's Tale* see <http://www.sd-editions.com/miller/index.html>. For further information on *The Nun's Priest's Tale* see <http://www.sd-editions.com/NP/index.html>.
6. Windram, Shaw, Robinson, and Howe "test the application of computer-based phylogenetic methods to the stemmatic analysis of manuscript relationships. Our results show that these methods […] are capable of producing stemmata in very close agreement with those produced by traditional stemmatic analysis" (2008: 443).

Works Cited

Barbrook, Adrian C., Christopher J. Howe, Norman Blake, and Peter Robinson. "The Phylogeny of *The Canterbury Tales*." *Nature* 394 (1998): 839. Print.

Bordalejo, Barbara. "The Transcription and Encoding of Pre-1520 *Cancionero* Manuscripts." *An Electronic Corpus of Fifteenth-Century Castilian* Cancionero *Manuscripts*. U of Liverpool, 2014. Web. 7 Mar. 2014. <http://cancionerovirtual.liv.ac.uk/bb-article/TransAndEnc.html>.

Cerquiglini, Bernard. *Éloge de la variante: Histoire critique de la philologie*. Paris: Éditions du Seuil, 1989. Print.

Chaucer, Geoffrey. *The Miller's Tale*. Ed. Peter Robinson. Birmingham: SD Editions, 2004. CD-ROM.

———. *The Nun's Priest's Tale*. Ed. Paul Thomas. Birmingham: SD Editions, 2006. CD-ROM.

COLLATE 2.0. Developed by Peter Robinson. Birmingham: SD Editions. 2003. Software program.

Dante Alighieri. *Commedia*. Ed. Prue Shaw. Birmingham: SD Editions, 2012. CD-ROM.

Dutton, Brian. *El cancionero del siglo XV, c. 1360–1520*. 7 vols. Salamanca: U de Salamanca, 1990–91. Print.

———. *Catálogo-Índice de la poesía cancioneril del siglo XV*. Madison: HSMS, 1982. Print.

An Electronic Corpus of Fifteenth-Century Castilian Cancionero *Manuscripts*. Dir. Dorothy Severin. U of Liverpool, 2007– .Web. 8 Jan. 2014. <http://cancionerovirtual.liv.ac.uk>.

Gallica. Bibliothèque nationale de France, 1997– . Web. 7 Mar. 2014. <gallica.bnf.fr>.

Junicode. Font. <http://www.junicode.sourceforge.net>.

Robinson, Peter. "About These Stemmata." *An Electronic Corpus of Fifteenth-Century Castilian* Cancionero *Manuscripts*. U of Liverpool, 2007. Web. 7 Mar. 2014. <http://cancionerovirtual.liv.ac.uk/pr-article/aboutstemmata.html>.

SD Editions. 2000– . Web. <http://www.sd-editions.com/>.

Windram, Heather F., Pure Shaw, Peter Robinson, and Christopher. J. Howe. "Dante's *Monarchia* as a Test Case for the Use of Phylogenetic Methods in Stemmatic Analysis." *Literary and Linguistic Computing* 23.4 (2008): 443–63. Print.